DOCUMENT RESUME

ED 450 150	TM 032 340
AUTHOR	Marsh, S. Neil
TITLE	Factor Scores, Structure Coefficients, and Communality Coefficients: It's All One General Linear Model.
PUB DATE	2001-02-00
NOTE	34p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (New Orleans, LA, February 1-3, 2001).
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Factor Analysis; Factor Structure; *Junior High School Students; Junior High Schools; *Scores
IDENTIFIERS	*Communality; *General Linear Model

ABSTRACT

This paper explains the meaning and use of three important factor analytic statistics: factor scores, factor structure coefficients, and communality coefficients. For the discussion, 301 observations of junior high school students 11 measured variables from a previous study are analyzed. While factors provide the researcher with general information, the factor scores are weighted combinations of scores on a series of measured variables. Four methods of calculating factor scores are discussed. These are: (1) the regression method; (2) the Bartlett method (M. Bartlett, 1937), which uses least squares procedures to minimize the sum of squares of the other unique factors over the range of variables; (3) the Anderson-Rubin method (T. Anderson and H. Rubin, 1956), which is similar to the Bartlett method with an added condition requiring that factors scores be orthogonal; and (4) the Thompson method (B. Thompson, 1993). Calculating factor structure coefficients and communality coefficients is described, and what each brings to the analysis is outlined. In the example studied, the interpretable results were 5 extracted factors and 301 factor scores for each of the factors, 55 structure coefficients, and 11 communality coefficients. The combined evaluation of these statistics enables the researcher to formulate interpretations more accurately. An attachment contains the syntax reference for the analysis. (Contains 9 tables and 23 references.) (SLD)

Running Head: GENERAL LINEAR MODEL

FACTOR SCORES, STRUCTURE COEFFICIENTS, AND COMMUNALITY COEFFICIENTS: IT'S ALL ONE GENERAL LINEAR MODEL

S. NEIL MARSH

TEXAS A&M UNIVERSITY

Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans, February 1-3, 2001.

Abstract

The purpose of factor analysis is to "summarize the interrelationships among the variables in a concise but accurate manner as an aid in conceptualization" (Gorsuch, 1983, p. 2). Kerlinger (1979) described factor analysis as "one of the most powerful methods yet for reducing variable complexity to greater simplicity" (p. 180). The purpose of the present paper is to explain the meaning and use of three important factor analytic statistics: factor scores, factor structure coefficients, and communality coefficients. In addition, four methods of calculating factor scores are discussed.

FACTOR SCORES, STRUCTURE COEFFICIENTS, AND COMMUNALITY COEFFICIENTS: IT'S ALL ONE GENERAL LINEAR MODEL

The purpose of factor analysis is to "summarize the interrelationships among the variables in a concise but accurate manner as an aid in conceptualization" (Gorsuch, 1983, p. 2). In attempting to find discrete yet meaningful insights about data (Horst, 1965), this summary should include the maximum amount of information from the original measured variables in as few latent or synthetic variables or factors as possible, so as to keep the solution parsimonious (Hetzel, 1996). Kerlinger (1979) described factor analysis as "one of the most powerful methods yet for reducing variable complexity to greater simplicity" (p. 180). Cattell (1978) wrote that factor analysis is "the furthest logical development and reigning queen of the correlation methods" (p. 4).

Simply, factor analysis, just like all other GLM analyses, looks at the relationship between measured variables and latent variables. However, some misuses and misconceptions about factor analysis can be attributed to the confusing language surrounding the method. As Pedhazur and Schmelkin (1991) wrote:

> Perusing even small segments of this [factor analysis] literature in an effort to understand what FA [factor

> analysis] is, how it is applied, and how the results are interpreted is bound to bewilder and frustrate most readers. This is due to a wide variety of contrasting and contradictory views on almost every aspect of FA, serious misconceptions surrounding it, and lack of uniformity in terminology and notation. (p. 590)

Garbarino (1996) also commented on the confusing language used to describe various parametric statistics:

> For example, we call the same systems of weights "equations" in regression, "factors" in factor analysis, "functions" or "rules" in discriminant analysis, and "functions" in canonical correlational analysis. We call the weights themselves "beta" weights in regression, "pattern coefficients" in factor analysis, and "standardized function coefficients" in discriminant analysis or canonical correlation analysis. The synthetic scores are called "yhat" in regression, "discriminant scores" in discriminant analysis, and "canonical function (or variate) scores" in canonical correlation analysis. (p. 3)

Thompson (1992) noted that various statistical concepts from different analyses (e.g., factor analysis, regression, canonical correlation analysis) "…are all analogous, but are given different names in different analyses…mainly to obfuscate the commonalities of [all] parametric methods, and to confuse graduate students" (pp. 906-907). Further, terms like "loadings" have been used ambiguously in referring to both factor structure and pattern coefficients (Thompson & Daniel, 1996).

The purpose of the present paper is to explain the meaning and use of three important factor analytic statistics: factor scores, factor structure coefficients, and communality coefficients. In making sense of factor analytic results, one must correctly identify and interpret these three sets of statistics (Wells, 1999). Actually, these results occur throughout analyses within the general linear model (Thompson, 2000), but are arbitrarily given different names. Despite the obvious importance of these parameters a number of articles have failed to interpret, as well as neglect to even report statistics like factor structure coefficients (Thompson, 1997).

Various ways of estimating factor scores will be compared and contrasted, including new non-centered estimation methods (Thompson, 1993). The paper will also illustrate that structure coefficients are bivariate correlation coefficients between the measured variables in the factor analysis with the factor scores yielded in the analysis (McMurray, 1987). Finally, communality coefficients (h^2) will be expressed as the R^2 between the factor scores on all the factors in a given solution with the scores on a given measured variable (Wells, 1999).

Heuristic Data

In providing a general overview of the computations and interpretations, as well as a step by step discussion of factor analysis, the Holzinger and Swineford (1939) data set will be used. For heuristic purposes, a portion of the original 301 student observations on 25 measured variables will be analyzed and present in a graphical format to assist in providing a concrete understanding of factor scores, communality coefficients, and factor structure coefficients. More specifically, a factor analysis using (a) cubes, simplification of Brigham's spatial relations test; (b) paper form board; (c) lozenges from Thorndike; (d) paragraph completion test; (e) word meaning test; (f) speeded addition test; (g) speeded code test; (h) speeded counting of dots; (i) memory of target words; (j) memory of target numbers; (k) memory of target shapes on all 301 junior high students was performed.

A factor analysis was completed and five factors were extracted. Although other types of factor extraction can be used, this paper applied principal components analysis, which always maintains (unlike other extraction methods)exactly the same correlations among the factor scores as well as between the factors. Factor analysis is just like all other General Linear Model statistics in that it is correlational (Thompson, 1991). Therefore, just like yhat scores in regression, factor scores are estimates of the latent constructs of primary interest to researchers.

To get factor scores, pattern coefficients, and structure coefficients we must first start with our raw data matrix. As can be seen in Table 1, gaining understandable information from a raw data matrix is virtually impossible. The same is true as regards the Table 2 z-score form of the data; Table 3 presents the variable labels. In fact, one important goal of factor analysis is finding and understanding existing relationships between observed variables as well as latent variables.

Therefore, rather than run analysis on the raw data matrix or the z score matrix, which may contain "random or unreliable information" (Horst, 1965, p. 469), factor analysis begins with an association matrix (e.g., correlation matrix, covariance matrix). Instead of using some other association matrix, the present factor analysis used a correlation matrix of the measured variables, which is special case of a variance-covariance matrix (Fan, 1996). Not only does the correlation matrix attempt to simplify the data, it also suggests existing relationships between measured variables.

In this paper, we will be especially interested in two Pearson correlation matrices. The Pearson correlation matrix of the measured variables presented in Table 4 is called "symmetric" because the number of columns and rows are equal. Table 5 presents the factor pattern coefficients "extracted" (Hetzel, 1996) from the Table 4 correlation matrix; the factor pattern coefficients are mathematically analogous to the beta weights derived in regression analysis. Table 6 presents the factor correlation coefficients.

The factor structure coefficients (unlike the pattern coefficients) are always correlation coefficients, as we shall see momentarily. Factor analysis structure coefficients are directly analogous to the structure coefficients derived throughout the general linear model, including regression (Thompson & Borrello, 1985) and canonical correlation analysis (Thompson, 2000).

In matrix algebra a matrix that when multiplied times another matrix yields that other matrix is called an "identity" matrix (i.e., the "identity matrix" is the matrix algebra equivalent to the number "1" in regular algebra). An identity matrix has 1's on the diagonal and 0's everywhere off the diagonal. Thus, the Table 6 factor correlation coefficients constitute an identity matrix.

Because the factor structure coefficient matrix (S_{VxF}) equals the factor pattern coefficient matrix (P_{VxF}) times the factor correlation matrix (R_{FxF}), and in present case $R_{FxF} = I_{FxF}$, here S_{VxF} equals P_{VxF}. This is why the numbers in Tables 5 and 7 are the same. Thus, the two matrices might have been presented as a single matrix, and labeled the "pattern/structure matrix."

Factor Scores

Wells (1999) pointed out that a number of people confuse factor scores with factors. Remember, a factor score matrix has "n" rows (one for every individual), while a factor matrix has "v" rows (as in a pattern/structure coefficient matrix). To understand the utility of factor scores we must first realize that factors are conceptual entities or latent variables. While factors provide the researcher with general information, the factor scores are detailed representations that attempt to help us understand these often confusing constructs.

Again, factor scores are similar to yhat scores in regression. In regression, the yhat scores provided a linear combination of an individual's score on a measured variable (Kachigan, 1982). Likewise, factor scores are weighted combinations of scores on a series of measured variables. A set of factor scores exists for every person on every component of a factor. Rather than being derived linearly, matrix algebra is used to estimate approximations of each individual's factor scores. Four possible methods (Gabarino, 1996; McDonald & Burr, 1967; Wells, 1999) calculate factor scores:

1. The Regression Method, which is available in SPSS, determines factor scores by multiplying the standardized score matrix by the inverse of the variable correlation matrix. Any matrix multiplied by an inverse matrix is actually dividing out or removing the presence of the inverse matrix. Therefore, the relationship between the variables is removed. Table 8 presents these factor scores for the present example.

$$F_{NxV} = Z_{NxV} R^{-1}{}_{VxV}$$

2. The Bartlett Method (also available in SPSS)uses least squares procedures to minimize the sum of squares of the unique factors over the range of variables (Bartlett, 1937). Because the sum of squares of the unique factors are minimized, non-common factors are used only to explain the

discrepancies between observed scores and those reproduced from the common factors. This method eventually leads to high correlation between factor scores and factors being estimated and ensures that the factor scores are the unbiased estimates.

3. The Anderson-Rubin Method (Anderson & Rubin, 1956) (also available in SPSS) is similar to the Bartlett Method with an added condition requiring that factor scores must be orthogonal. The resulting equation is more complex and produces factor estimates whose correlations form an identity matrix. These estimates are neither "univocal" nor unbiased but do have reasonably high correlations with the factors.

4. The Thompson Method (Thompson, 1993)could easily be performed with SPSS using syntax creates factor scores that are not generated in z score form. Although, the standard deviation of the standardized factor scores is 1, as in z score based formulas, the means of the measured variables are added back into the factor scores (making them "non-centered"). Therefore, researchers can compare the mean factor scores across factors within a given analysis. Basically, this method requires that the researcher convert variables into standardized form, add original variable means back onto the standardized estimates, and finally,

obtain factor scores by multiplying these non-centered but standardized values by the weight matrix ($W_{VxF} = P_{VxF} R^{-1}_{VxV}$)

$$(W_{VxF} = P_{VxF} R^{-1}_{VxV}) * Z_{NxV} = F_{NxF}$$

Factor Structure Coefficients

If the researcher wants to know the importance of a variable to a specific factor in the presence of the other variables the factor pattern coefficients or weights must be consulted (especially if factors were correlated, as in an oblique rotation). However, if we are interested in the bivariate relationship between a measured variable and specific latent variable, we look at structure coefficients (Thompson, 1997). Not only are structure coefficients essential in the interpretation of univariate statistics (e.g., multiple regression), interpretation of multivariate statistics often calls for separate assessment of structure coefficients apart from weights (Thompson, 1992). Because the five factors discerned in this paper were extracted using principal components analysis and a varimax rotation, the extracted factors are orthogonal (i.e., uncorrelated). Wells (1999) pointed out that factors are "always perfectly uncorrelated upon initial extraction, and remain uncorrelated if an 'orthogonal' rotation is used" (p. 126).

The fifth factor was extracted for comparative purposes and would not normally be extracted due to the failure to met extraction criteria (e.g., scree plot test, eigenvalue test). In this factor analysis, the structure coefficients have the same value as the pattern coefficients. Better yet, the rotated pattern coefficient matrix and factor structure coefficient matrices are equal because the factor correlation matrix is an identity matrix.

In the case of factor analysis, we are interested in the relationship of specific measured variables with specific factors or latent constructs (Gabarino, 1996). For example, the measured variable t2 has a factor structure coefficient of (.883) with Factor 4. Because structure coefficients are a "score world" statistic we must "square to compare". Therefore, the squared structure coefficient, or the estimate of the bivariate relationship between t2 and Factor 4 can be expressed the "area world" form as .780 (.883 x .883 = .780). Squared structure coefficients provide us additional methods of interpretations that are very useful to factor analysis: (1) Eigenvalues; (2) Communality coefficients. While a discussion on eigenvalues might be beneficial, it is beyond the scope of this paper. However, Stevens (1996) provides an effective treatment of the topic.

Communality Coefficients (h^2)

Whereas structure coefficients look at bivariate relationships between measured variables and a single factor, communality coefficients look at or estimate variance accounted for between one measured variable across all the factors. Communality coefficients are best described as each variable's variance that has been reproduced by the extracted factors (Gorsuch, 1985).

The communality coefficient is the sum of squared structure coefficients across all extracted factors. For example, if communality of variable t2 were 1.00 or 100%, we could entertain the idea that all of the variance in t2 was accounted for by the factors or all of the variance in t2 was useful in identifying the factors. Communality coefficients can only be positive or zero values, because these estimates are dealing with squared numbers. McMurray (1987) pointed out that communality coefficients can be considered the multiple correlation coefficients in factor analysis.

This is best seen in each of the variable's multiple R squared. In this writing, a regression with each of the variables being the dependent measure and the factor scores being the predictors was performed. Each R squared value is equal to the variable's communality coefficient; see Table

9. It is taking into account all of the predictors' influence or accounting for the variance of the factor scores. Essentially, factors are reproducing the reliable or usable variance/information from each measured variable. Further, if we sum all of the communality coefficients and divide by the number of variables, we see the total proportion of variance accounted for by all five factors equals 72.158.

In addition, communalities are also termed as the proportion of variance that was useful in identifying or delineating the factors (Gorsuch, 1983). Communality coefficients are also lower-bound estimates of reliability. In principal components analysis (used in this paper) the initial reliability estimate of each variable is considered perfect or 100% reliable. Obviously, it is highly unlikely that all the measured variables will actually be found to be completely reliable. Therefore, other types of factor extraction do not assume perfect reliability.

Conclusion

In our present study of the eleven measured variables from the 301 Holizinger (1939) students the interpretable results would be (1) five extracted factors and 301 factor scores for each of the factors;(2) 55 structure coefficients would exist, one for each measured variable and its relationship with each of the extracted factors, and (3) 11 communality coefficients could be determined, one for each measured variable. As in other GLM procedures, the combined evaluation of factor scores, structure coefficients, pattern coefficients, and communality coefficients enables the researcher to more accurately formulate interpretations based on the results.

References

Anderson, T. W., & Rubin, H. (1956). Statistical inference in factor analysis. Proceedings of Third Berkeley Symposium of Mathematical Statistics and Probability, 5, 111-150.

Bartlett, M. S. (1937). The statistical conception of mental factors. British Journal of Psychology, 28, 97-104.

Cattell, R. B. (1978). The scientific use of factors analysis in behavioral and life sciences. New York: Plenum.

Fan, X. (1996). Canonical correlation analysis as a general analytic model. In B. Thompson (Ed.), Advances in social science methodology (Vol. 4, pp. 71-94). Greenwich, CT: JAI Press.

Garbarino, J. J. (1996, January). Alternate ways of computing factor scores. Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED 395 042)

Gorsuch, R. L. (1983). Factor analysis (2nd ed.). Hilldale, NJ: Erlbaum.

Hetzel, R. D. (1996). A primer on factor analysis with comments on analysis and interpretation patterns. In

B. Thompson (Ed.), Advances in social science methodology (Vol. 4, pp. 175-206). Greenwich, CT: JAI Press.

Holzinger, K. J., & Swineford, F. (1939). A study in factor analysis: The stability of bi-factor solution (No. 48). Chicago: University of Chicago. (data on pp. 81-91)

Horst,P. (1965). Factor analysis of data matrices. New York: Holt, Rinehart, & Winston.

Kachigan,S. K. (1982). Multivariate statistical analysis. New York: Radius Press.

Kerlinger,F. N. (1979). Behavioral research: A conceptual approach. New York: Holt, Rinehart, & Winston.

McDonald, R. P., & Burr, E. J. (1967). A comparison of four methods of constructing factor scores. Psychometrika, 32, 381-401.

McMurray, M. A. (1987, January). The concepts underlying structure coefficients, communality, and factor scores in the exploratory factor analytic case. Paper presented at the annual meeting of the Southwest Educational Research Association, Dallas, TX. (ERIC Document Reproduction Service No. ED 284 873)

Pedhazur,E. J.,& Schmelkin, L. P. (1991). Measurement, design, and analysis: An integrated approach. Hillsdale, NJ: Erlbaum.

Stevens, J. (1996). Applied multivariate statistics for social sciences (3rd ed.). Mahwah, NJ: Erlbaum.

Thompson, B. (1991). A primer on the logic and use of canonical correlation analysis. Measurement and Evaluation in Counseling and Development, 24, 80-95.

Thompson, B. (1992). DISCSTRA: A computer program that computes bootstrap resampling estimates of descriptive discriminant analysis function and structure coefficients and group centroids. Educational and Psychological Measurement, 52, 905-911.

Thompson, B. (1993). Calculation of standardized, noncentered factor scores: An alternative to conventional factor scores. Perceptual and Motor Skills, 77, 1128-1130.

Thompson, B. (1997). The importance of structure coefficients in structural equation modeling confirmatory analysis. Educational and Psychological Measurement, 57, 5-19.

Thompson, B. (2000). Canonical correlation analysis. In L. Grimm & P. yarnold (Eds.), Reading and understanding more multivariate statistics (pp. 285-316). Washington, DC: American Psychological Association.

Thompson, B., & Borrello, G. M. (1985). The importance of structure coefficients in regression research. Educational and Psychological Measurement, 45, 203-209.

Thompson, B. & Daniel, L.G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. Educational and Psychological Measurement, 56, 197-208.

Wells, R. D. (1999). Factor scores and factor structure and communality coefficients. In B. Thompson (Ed.), Advances in Social Science Methodology (Vol. 5, pp. 123-138). Greenwich, CT: JAI Press.

Table 1: Raw Data Matrix

	Variable									
Case	T2	T3	T4	T6	T9	T10	T11	T12	T14	T15	T16
1	31	12	3	7	9	78	74	115	170	86	96
2	21	12	17	5	9	87	84	125	184	85	100
3	21	12	15	3	3	75	49	78	170	85	95
4	31	16	24	8	17	69	65	106	181	80	91
5	19	12	7	8	18	85	63	126	187	99	104
6	20	18	11	3	6	100	92	133	164	84	104
7	24	12	8	10	20	108	65	124	121	71	78
8	25	13	15	11	9	78	80	103	184	95	106
9	23	11	12	8	19	104	52	93	184	91	105
10	21	6	10	8	18	95	74	91	175	92	100
11	23	13	16	6	11	86	60	114	173	86	107
12	24	15	23	8	19	85	71	103	167	103	108
13	18	15	33	8	16	135	68	104	166	92	103
14	22	16	14	14	11	118	68	94	186	86	102
15	23	13	29	15	21	92	64	87	168	84	102
17	19	18	12	6	9	85	58	133	164	84	100
18	27	17	18	6	12	92	61	105	149	88	98
19	21	14	32	13	26	90	94	97	171	74	107
20	35	13	24	11	18	80	62	107	186	99	105
21	32	17	16	5	4	60	44	92	169	96	96
22	34	16	33	6	15	103	73	109	169	94	105
23	25	16	15	10	22	80	62	92	166	83	101
24	22	15	13	8	7	134	64	106	173	96	99
25	22	12	10	7	9	108	60	92	154	92	106

Table 2: Standard Score Matrix (13 cases)

	Variable										
Case	zt2	zt3	zt4	zt6	zt9	zt10	zt11	zt12	zt14	zt15	zt16
1	1.41	-.79	-1.66	-.62	-.82	-.73	.31	.22	-.45	-.52	-.015
2	-.72	-.79	-.11	-1.20	-.82	-.37	.95	.71	.-77	-.65	-.33
3	-.71	-.79	-.33	-1.77	-1.60	-.85	-1.29	-1.61	-.45	-.65	-.99
4	1.41	.63	.66	-.34	.22	1.09	-.27	-.22	.51	-1.29	-1.51
5	-1 .14	-.7	.9	-1.22	-.34	-.45	-.39	.76	1.03	1.16	.19
6	-.92	-1.14	.00	-1.77	-1.21	.15	1.46	1.11	-.97	-.78	.19
7	-.07	-.79	-1.10	.23	.61	.47	-.27	.66	-4.70	-2.46	-3.21
8	.14	-.43	-.33	.52	-.82	-.73	.69	-.37	.77	.65	.46
9	-.29	-1.14	-.66	-.34	.48	.31	-1.10	-.87	.77	.13	.33
10	-.71	-1.49	-1.33	-.34	.35	-.05	.31	-.96	-.01	.26	-.33
11	-.29	-.43	-.22	-.91	-.56	-.41	-.58	.17	-.19	-.52	.59
12	-.07	.27	.55	-.34	.48	-.45	.12	-.37	-.711	.68	.72
13	-1.35	.27	1.66	-.34	.09	1.55	-.07	-.32	-.79	.26	.06

Table 3: Variable Labels

Variable	Labels
t2	CUBES, SIMPLIFICATION OF BRIGHAM'S SPATIAL RELATIONS TEST
t3	PAPER FORM BOARD-SHAPES THAT CAN BE COMBINED TO FORM A TARGET
t4	LOZENGES FROM THORNDIKE-SHAPES FLIPPED OVER THEN IDENITIFY TARGET
t6	PARAGRAPH COMPREHENSION TEST
t9	WORD MEANING TEST
t10	SPEEDED ADDITION TEST
t11	SPEEDED CODE TEST
t12	SPEEDED COUNTING OF DOTS
t14	MEMORY OF TARGET WORDS
t15	MEMORY OF TARGET NUMBERS
t16	MEMORY OF TARGET SHAPES

Table 4:
Correlation Matrix

Variable	t2	t3	t4	t6	t9	t10	t11
t2	1.000						
t3	0.238	1.000					
t4	0.340	0.305	1.000				
t6	0.153	0.212	0.159	1.000			
t9	0.193	0.239	0.198	0.704	1.000		
t10	-0.076	0.040	0.072	0.174	0.121	1.000	
t11	0.108	0.126	0.199	0.342	0.290	0.447	1.000
t12	0.092	0.177	0.186	0.107	0.150	0.487	0.398
t14	0.068	0.073	0.128	0.222	0.172	0.093	0.225
t15	0.085	0.036	0.212	0.069	0.052	0.109	0.140
t16	0.236	0.184	0.305	0.241	0.253	0.117	0.305

Table 5: Rotated pattern
and structure matrix ext=p.c.a./rotation=varimax)

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
t2	-0.042	0.131	0.007	0.883	0.001
t3	0.055	0.164	0.033	0.149	0.933
t4	0.149	0.023	0.219	0.632	0.355
t6	0.117	0.897	0.111	0.047	0.071
t9	0.093	0.897	0.06	0.126	0.119
t10	0.838	0.065	0.075	-0.139	-0.016
t11	0.677	0.327	0.207	0.139	-0.076
t12	0.807	-0.022	-0.025	0.135	0.181
t14	0.015	0.193	0.797	-0.072	0.019
t15	0.072	-0.106	0.792	0.077	0.02
t16	0.139	0.2	0.623	0.344	0.054

Table 6: factor score (created using Regression) correlation matrix

	FSCORE1	FSCORE2	FSCORE3	FSCORE4	FSCORE5
FSCORE1	1.00	0.00	0.00	0.00	0.00
FSCORE2	0.00	1.00	0.00	0.00	0.00
FSCORE3	0.00	0.00	1.00	0.00	0.00
FSCORE4	0.00	0.00	0.00	1.00	0.00
FSCORE5	0.00	0.00	0.00	0.00	1.00

Table 7 :
Structure
coefficients

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
t2	-0.042	0.131	0.007	0.883	0.001
t3	0.055	0.164	0.033	0.149	0.933
t4	0.149	0.023	0.219	0.632	0.355
t6	0.117	0.897	0.111	0.047	0.071
t9	0.093	0.897	0.06	0.126	0.119
t10	0.838	0.065	0.075	-0.139	-0.016
t11	0.677	0.327	0.207	0.139	-0.076
t12	0.807	-0.022	-0.025	0.135	0.181
t14	0.015	0.193	0.797	-0.072	0.019
t15	0.072	-0.106	0.792	0.077	0.02
t16	0.139	0.2	0.623	0.344	0.054

Table 8:
Communality
coefficients

Variables	Communality	Rsquared
t2	0.798	0.798
t3	0.924	0.924
t4	0.596	0.596
t6	0.839	0.839
t9	0.814	0.814
t10	0.732	0.732
t11	0.633	0.633
t12	0.704	0.704
t14	0.679	0.679
t15	0.649	0.649
t16	0.568	0.568

Table 9: Abbreviated list of 25 individuals' factor scores

Case	FSCORE1	FSCORE2	FSCORE3	FSCORE4	FSCORE5
1	-.09121	-.44487	-.90708	.67799	-1.39301
2	.61821	-.88595	.10662	-.35552	-.60147
3	-1.27638	-1.52936	-.42944	-.47816	-.38047
4	-.80970	.11520	-1.01208	1.06929	.54899
5	-.05228	-.09570	1.17991	-1.33814	-.48943
6	1.46515	-1.48908	-.61697	.06798	-.98865
7	.88869	.60172	-4.77595	.00761	-.75787
8	-.34209	.00025	.94162	.04480	-.70099
9	-.66314	.18922	.63056	-.55803	-1.03886
10	-.30151	.32486	.18245	-.90708	-1.64946
11	-.18213	-.72998	-.07943	.04028	-.23886
12	-.33370	-.22387	.76709	.39468	.30503
13	.75589	-.47021	-.00926	-.44059	.86327
14	-.11508	.68688	.29098	-1.11849	.50387
15	-.72251	1.37653	-.62559	.27270	-.25045
16	-.88177	-.46100	-.16702	-.93657	-.09645
17	.23063	-1.00112	-.92091	-1.00163	1.78356
18	-.18653	-.82008	-1.46503	.67876	.91364
19	.12848	1.72671	-.91890	.35857	-.09471
20	-.78864	.26088	.88273	1.86228	-.80601
21	-1.62734	-1.54124	-.14411	1.07094	.87014
22	.17007	-.81524	-.07318	2.30952	.33252
23	-.92434	.82994	-.93307	-.00363	.48972
24	.59265	-.90309	.27003	-.92394	.32785
25	-.12663	-.72006	-.43979	-.27037	-.83780

SYNTAX REFERENCE

```
SET BLANKS=SYSMIS UNDEFINED=WARN PRINTBACK=LISTING.
DATA LIST
      FILE='c:\WINDOWS\Desktop\New Folder\holzinger.dta' FIXED
RECORDS=2 TABLE /1
      Id 1-3 sex 4-4 ageyr 6-7 agemo 8-9 t1 11-12 t2 14-15 t3 17-18 t4
20-21 t5 23-24 t6 26-27 t7 29-30 t8 32-33 t9 35-36 t10 38-40 t11 42-44
t12 46-48 t13 50-52 t14 54-56 t15 58-60 t16 62-64 t17 66-67 t18 69-70
t19 72-73 t20 74-76 t21 78-79 /2 t22 11-12 t23 14-15 t24 17-18 t25 20-
21 t26 23-24 .
EXECUTE.
COMPUTE SCHOOL=1.
IF (id GT 200) SCHOOL=2.
IF (id GE 1 AND id LE 85) GRADE=7.
IF (id GE 86 AND id LE 168) GRADE=8.
IF (id GE 201 AND id LE 281)GRADE=7.
IF (id GE 282 AND id LE 351)GRADE=8.
IF (id GE 1 AND id LE 44)TRACK=2.
IF (id GE 45 AND id LE 85)TRACK=1.
IF (id GE 86 AND id LE 129)TRACK=2.
IF (id GE 130) TRACK=1.
PRINT FORMATS SCHOOL TO TRACK(F1.0).
VALUE LABLES SCHOOL(1)PASTEUR (2)GRANT-WHITE/TRACK (1)JUNE PROMTIONS(2)
FEB PROMOTIONS/.
VARIABLE LABELS t1 VISUAL PERCEPTION TEST FROM SPEARMAN VPT, PART III
t2 CUBES, SIMPLIFICATION OF BRIGHAM'S SPATIAL RELATIONS
t3 PAPER FORM BORAD-SHAPES THAT CAN BE COMBINED TO FORM
t4 LOZENGES FROM THORNDIKE-SHAPES FLIPPED OVER THEN IDENT
t5 GENERAL INFORMATION VERBAL TEST
t6 PARAGRAPH COMPREHENSION TEST
t7 SENTENCE COMPLETION TEST
t8 WORD CLASSIFICATION-WORD THAT DOES NOT BELONG
t9 WORD MEANING TEST
t10 SPEEDED ADDITION
t11 SPEEDED CODE TEST
t12   SPEEDED COUNTING OF DOTS
t13 SPEEDED DISCRIM STRAIGHT AND CURVE
t14 MEMORY OF TARGET WORDS
t15 MEMORY OF TARGET NUMBERS
t16 MEMORY OF TARGET SHAPES
t17 MEMORY OF OBJECT-NUMBER ASSOCIATION
t18 MEMORY OF NUMBER-OBJECT ASSOCIATION
t19 MEMORY OF FIGURE-WORD ASSOCIATION
t20 DEDUCTIVE MATH ABILITY
t21 MATH NUMBER PUZZLES
t22 MATH WORD PROBLEM
t23 COMPLETION OF A MATH NUMBER SERIES
t24 WOODY-MCCALL MIXED MATH FUND
t25 REVISION OF T3-PAPER FORM
t26  FLAGS-POSSIBLE SUBSTITUTE FOR T4 LOZENGES .
SUBTITLE 'FACTOR 11 VARIABLES INTO FACTORS***'.
```

```
FACTOR VARIABLES=t2 t3 t4 t6 t9 t10 t11 t12 t14 t15
t16/PRINT=ALL/PLOT=EIGEN/CRITERIA=FACTORS(5)/EXTRACTION=PC/ROTATION=VAR
IMAX/SAVE=REG(ALL FSCORE).
LIST VARIABLES=FSCORE1 FSCORE2 FSCORE3 FSCORE4
FSCORE5/CASES=25/FORMAT=NUMBERED.
Subtitle '1a. Factor scores BARTLETT method****'.
FACTOR VARIABLES=t2 t3 t4 t6 t9 t10 t11 t12 t14 t15
t16/CRITERIA=FACTORS (5)/EXTRACTION=PC/ROTATION=VARIMAX/SAVE=BART (ALL
FSCOR).
VARIABLE LABELS FSCORE1 'SPEED bart'
FSCORE2 'MEMORY bart' FSCORE3 FSCORE4 FSCORE5.
Subtitle '1b. Factor scores Anderson-Rubin method****'.
FACTOR VARIABLES=t2 t3 t4 t6 t9 t10 t11 t12 t14 t15
t16/CRITERIA=FACTORS (5)/EXTRACTION=PC/ROTATION=VARIMAX/SAVE=BART (ALL
FSCR).
VARIABLE LABELS FSCR1 'SPEED ar'
FSCR2 'MEMORY ar' FSCR3 FSCR4 FSCR5.
Subtitle '2a. Compute z-score****'.
DESCRIPTIVES VARIABLES=t2 t3 t4 t6 t9 t10 t11 t12 t14 t15 t16/save.
print formats zt2 zt3 zt4 zt6 zt9 zt10 zt11 zt12 zt14 zt15 zt16(F8.5) .
List variables=zt2 zt3 zt4 zt6 zt9 zt10 zt11 zt12 zt14 zt15
zt16/cases=25.
Subtitle '2b. Prove z-scores are z-scores***'.
DESCRIPTIVES VARIABLES=zt2 zt3 zt4 zt6 zt9 zt10 zt11 zt12 zt14 zt15
zt16.
Subtitle '2c. Compute regression factor scores hard way***'.
COMPUTE FSHARD1=(-.042*zt2)+ (.055
*zt3)+(.149*zt4)+(.117*zt6)+(.093*zt9)+
(.838*zt10)+(.677*zt11)+(.807*zt12) +(.015*zt14)+(.072*zt15)+(.139 *
zt16).
COMPUTE FSHARD2=(.131*zt2)+
(.164*zt3)+(.023*zt4)+(.897*zt6)+(.897*zt9)+ (.065*zt10)+(.327
*zt11)+(-.022 *zt12) +(.193*zt14)+(-.106*zt15)+(.2* zt16).
COMPUTE FSHARD3=(.007*zt2)+ (.033*zt3)+(.219
*zt4)+(.111*zt6)+(.06*zt9)+ (.075*zt10)+(.207*zt11)+(-.025*zt12)
+(.797*zt14)+(.792 *zt15)+(.623*zt16).
COMPUTE FSHARD4=(.883*zt2)+ (.149*zt3)+(.632*zt4)+(.047 *zt6)+(.126
*zt9)+ (-.139*zt10)+(.139 *zt11)+(.135 *zt12) +(-
.072*zt14)+(.077*zt15)+(.344*zt16).
COMPUTE FSHARD5=(.001*zt2)+
(.933*zt3)+(.355*zt4)+(.071*zt6)+(.119*zt9)+ (-.016*zt10)+(-
.076*zt11)+(.181*zt12) +(.019*zt14)+(.02 *zt15)+(.054*zt16) .
VARIABLE LABELS FSHARD1 'SPEED hard'
FSHARD2  'MEMORY hard' FSHARD3 FSHARD4 FSHARD5.
Subtitle '3a. Compute Thompson factor scores ***'.
COMPUTE TT2=zt2+24.35 .
COMPUTE TT3=zt3+14.23.
COMPUTE  TT4=zt4+18.00.
COMPUTE TT6=zt6+9.18.
COMPUTE TT9=zt9+15.30.
COMPUTE TT10=zt10+96.28 .
COMPUTE TT11=zt11+69.16.
COMPUTE  TT12=zt12+ 110.54.
COMPUTE TT14=zt14+175.15 .
COMPUTE TT15=zt15+90.01 .
COMPUTE TT16=zt16+102.52.
```

```
COMPUTE FSBT1=(-.042*zt2)+ (.055*zt3)+(.149*zt4)+(.117*zt6)+(.093*zt9)+
(.838*zt10)+(.677*zt11)+(.807*zt12) +(.015*zt14)+(.072*zt15)+(.139 *
zt16).
COMPUTE FSBT2=(.131*zt2)+ (.164*zt3)+(.023*zt4)+(.897*zt6)+(.897*zt9)+
(.065*zt10)+(.327 *zt11)+(-.022 *zt12) +(.193*zt14)+(-.106*zt15)+(.2*
zt16).
COMPUTE FSBT3=(.007*zt2)+ (.033*zt3)+(.219 *zt4)+(.111*zt6)+(.06*zt9)+
(.075*zt10)+(.207*zt11)+(-.025*zt12) +(.797*zt14)+(.792
*zt15)+(.623*zt16).
COMPUTE FSBT4=(.883*zt2)+ (.149*zt3)+(.632*zt4)+(.047*zt6)+(.126 *zt9)+
(-.139*zt10)+(.139 *zt11)+(.135 *zt12) +(-
.072*zt14)+(.077*zt15)+(.344*zt16).
COMPUTE FSBT5=(.001*zt2)+ (.933*zt3)+(.355*zt4)+(.071*zt6)+(.119*zt9)+
(-.016*zt10)+(-.076*zt11)+(.181*zt12) +(.019*zt14)+(.02
*zt15)+(.054*zt16) .
VARIABLE LABELS FSBT1 'SPEED thompson'
 FSBT2 'MEMORY thompson'  FSBT3 FSBT4 FSBT5.
Subtitle '4. Show factor score relationship ***'.
LIST VARIABLES=FSCORE1 FSCORE2 FSCORE3 FSCORE4 FSCORE5 FSCR1 FSCR2
FSCR3 FSCR4 FSCR5 FSHARD1 FSHARD2 FSHARD3 FSHARD4 FSHARD5
FSBT1 FSBT2 FSBT3 FSBT4 FSBT5/CASES=25.
DESCRIPTIVES  VARIABLES=FSCORE1 FSCORE2 FSCORE3 FSCORE4 FSCORE5 FSCR1
FSCR2 FSCR3 FSCR4 FSCR5 FSHARD1 FSHARD2 FSHARD3 FSHARD4 FSHARD5
FSBT1 FSBT2 FSBT3 FSBT4 FSBT5 .
CORRELATIONS VARIABLES=FSCORE1 FSCORE2 FSCORE3 FSCORE4 FSCORE5 FSCR1
FSCR2 FSCR3 FSCR4 FSCR5 FSHARD1 FSHARD2 FSHARD3 FSHARD4 FSHARD5
FSBT1 FSBT2 FSBT3 FSBT4 FSBT5 .
CORRELATION VARIABLES=t2 t3 t4 t6 t9 t10 t11 t12 t14 t15 t16 WITH
FSCORE1 FSCORE2 FSCORE3 FSCORE4 FSCORE5 .
CORRELATION VARIABLES=FSCORE1 FSCORE2 FSCORE3 FSCORE4 FSCORE5.
COMMENT 'REGRESSION WITH EACH DEPENDENT VARIABLE BEING A DEP VAR WITH
THE OTHER DEP VARIABLES AND FSCORES BEING PREDICTOR VARIABLES'.
COMMENT ' MULTIPLE R-SQUARE IN REGRESSION EQUALS THE COMMUNALITY
COEFFICIENT FOR EACH MEASURED VARIABLE'.
Regression Variables=FSCORE1 FSCORE2 FSCORE3 FSCORE4 FSCORE5 t2 t3 t4
t6 t9 t10 t11 t12 t14 t15 t16/dependent=t2/enter FSCORE1 FSCORE2
FSCORE3 FSCORE4 FSCORE5.
Regression Variables=FSCORE1 FSCORE2 FSCORE3 FSCORE4 FSCORE5 t2 t3 t4
t6 t9 t10 t11 t12 t14 t15 t16/dependent=t3/enter FSCORE1 FSCORE2
FSCORE3 FSCORE4 FSCORE5.
Regression Variables=FSCORE1 FSCORE2 FSCORE3 FSCORE4 FSCORE5 t2 t3 t4
t6 t9 t10 t11 t12 t14 t15 t16/dependent=t4/enter FSCORE1 FSCORE2
FSCORE3 FSCORE4 FSCORE5.
Regression Variables=FSCORE1 FSCORE2 FSCORE3 FSCORE4 FSCORE5 t2 t3 t4
t6 t9 t10 t11 t12 t14 t15 t16/dependent=t6/enter FSCORE1 FSCORE2
FSCORE3 FSCORE4 FSCORE5.
Regression Variables=FSCORE1 FSCORE2 FSCORE3 FSCORE4 FSCORE5 t2 t3 t4
t6 t9 t10 t11 t12 t14 t15 t16/dependent=t9/enter FSCORE1 FSCORE2
FSCORE3 FSCORE4 FSCORE5.
Regression Variables=FSCORE1 FSCORE2 FSCORE3 FSCORE4 FSCORE5 t2 t3 t4
t6 t9 t10 t11 t12 t14 t15 t16/dependent=t10/enter FSCORE1 FSCORE2
FSCORE3 FSCORE4 FSCORE5.
Regression Variables=FSCORE1 FSCORE2 FSCORE3 FSCORE4 FSCORE5 t2 t3 t4
t6 t9 t10 t11 t12 t14 t15 t16/dependent=t11/enter FSCORE1 FSCORE2
FSCORE3 FSCORE4 FSCORE5.
```

```
Regression Variables=FSCORE1 FSCORE2 FSCORE3 FSCORE4 FSCORE5 t2 t3 t4
t6 t9 t10 t11 t12 t14 t15 t16/dependent=t12/enter FSCORE1 FSCORE2
FSCORE3 FSCORE4 FSCORE5.
Regression Variables=FSCORE1 FSCORE2 FSCORE3 FSCORE4 FSCORE5 t2 t3 t4
t6 t9 t10 t11 t12 t14 t15 t16/dependent=14/enter FSCORE1 FSCORE2
FSCORE3 FSCORE4 FSCORE5.
Regression Variables=FSCORE1 FSCORE2 FSCORE3 FSCORE4 FSCORE5 t2 t3 t4
t6 t9 t10 t11 t12 t14 t15 t16/dependent=15/enter FSCORE1 FSCORE2
FSCORE3 FSCORE4 FSCORE5.
Regression Variables=FSCORE1 FSCORE2 FSCORE3 FSCORE4 FSCORE5 t2 t3 t4
t6 t9 t10 t11 t12 t14 t15 t16/dependent=t16/enter FSCORE1 FSCORE2
FSCORE3 FSCORE4 FSCORE5.
```